Pace Law Review

Volume 37
Issue 1 *Fall 2016*

Article 1

March 2017

On the Categorical Approach to Free Speech – And the Protracted Failure to Delimit the True Threats Exception to the First Amendment

Wayne Batchis
University of Delaware, batchisw@udel.edu

On the Categorical Approach to Free Speech – And the Protracted Failure to Delimit the True Threats Exception to the First Amendment

Dr. Wayne Batchis

Introduction

On June 1, 2015, the Supreme Court decided *Elonis v. United States* on statutory rather than constitutional grounds.[1] In doing so, it turned away an important opportunity to provide needed clarification of true threats, a category of expression relegated to a lower level of protection by the Court almost a half-century ago.[2] The categorical approach to free speech made its first explicit appearance in Supreme Court case law in 1942.[3] Since that time, the Court has relied heavily on this method of constitutional interpretation, carving out discrete exceptions from the seemingly absolutist mandate of the First Amendment that Congress make no law abridging the freedom of speech.[4] Although the categorical approach – frequently front and center in First Amendment adjudication – has been with us for almost seventy-five years, it rests on a surprisingly unsettled theoretical foundation. It is an indispensable doctrinal tool with a puzzling and sometimes contradictory array of justifications and operating instructions. In this piece, I attempt to clear up the confusion. I examine the evolution of

* Department of Political Science and International Relations, University of Delaware, 459 Smith Hall Newark, DE 19716-2574. batchisw@udel.edu. Work: (302)831-1934. Cell: (215)410-1173.

1. Elonis v. United States, 135 S. Ct. 2001 (2015).
2. *See* Watts v. United States, 394 U.S. 705, 708 (1969).
3. *See* Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).
4. U.S. CONST. amend. I.

this approach to the First Amendment. I critically assess the famous dictum from *Chaplinsky v. New Hampshire* that is responsible for establishing this system of classification. I scrutinize a number of possible interpretations of *Chaplinsky* and explore the disparate scholarly and judicial perspectives on this mode of constitutional interpretation. Finally, I move from the foundations and justifications of the categorical approach to the way this system works in practice. I argue that if the Court is to maintain its fidelity to an effectual categorical system of First Amendment adjudication – one that is properly respectful of the high stakes for free expression and democratic self-governance – it is vitally important that Court adequately define and operationalize respective categories. The final third of this article delves into one such category: true threats. I closely examine the Court's true threats jurisprudence and look critically at the recent *Elonis* decision, contrasting the Court's protracted failure to define and delimit true threats with the comparatively robust guidance it has offered with other discrete categories.

Content-based discrimination has long been understood as one of the most troublesome forms of speech suppression. Unlike mere restrictions on how or where views may be expressed, targeting particular ideas because of their substance and penalizing or prohibiting them evokes draconian images of political oppression by tyrants, Orwellian mind control, and raw censorship. Yet, although the First Amendment is written in absolutist language, it has never been understood to offer absolute protection. There has, quite simply, always been a wide array of expression – from fraud to slander, from child pornography to coercive threats – that overwhelming majorities feel is intolerable and inconsistent with an ordered civil society. The Supreme Court has settled upon a range of doctrinal approaches to free speech that seek to both accommodate this reality, yet pay proper heed to a fundamental constitutional right.

One dominant methodology, utilized in most cases where expression is regulated based upon its content, is referred to as the categorical approach. In some respects, it would seem to offer the best of all worlds: it allows for a default rule that – consistent with the unequivocal language of the First

Amendment itself – affords virtually absolute protection against content based suppression, yet at the same time permits discrete exceptions to be carved out when the abridgement applies to certain narrow types of ostensibly "low value" or especially harmful speech.[5] While the categorical approach certainly has its critics, many free speech advocates applaud this method because it seems to minimize the risk of ad hoc balancing, a case-by-case approach that makes all speech susceptible to suppression if, in toto, a judge happens to determine that the circumstances weigh on the speech-restrictive side of the scale.[6]

A categorical approach, however, is only as good as the categories that comprise it. Indeed, a categorical system with insufficiently defined categories may ultimately backfire. It may be *less* protective than a wholesale balancing system. At the same time, it may provide the misleading appearance of a Court willing to place constraints on its own authority in the interest of principled First Amendment jurisprudence. On its face, a categorical structure establishes a consistent rule-based methodology that constrains courts, promoting – if not ensuring – disciplined predictability in free speech jurisprudence; but when the Supreme Court holds tight to its most important card – the definition and doctrinal mechanics of the category itself – it does the very opposite. It invites courts and legislatures to construe an entire category of expression as broadly or idiosyncratically as it likes. Until a higher court intervenes (or in the case of legislatures, any court with jurisdiction), the default is a far-reaching and malleable lack of First Amendment protection. Today, with the rapidly expanding landscape of electronic idea conveyance, quelling free speech uncertainty and insecurity is arguably more vitally important than ever. Yet, in the 2015 decision of *Elonis v. United States*, a Court notable for its highly speech-protective jurisprudence willfully rejected one of its most high-profile and potentially impactful opportunities to do just this.[7] Even more troubling, it did so with regard to a highly consequential

5. GEOFFREY R. STONE et al., CONSTITUTIONAL LAW 1134 (7th ed. 2013).

6. *See, e.g.*, THOMAS I. EMERSON, TOWARD A GENERAL THEORY OF THE FIRST AMENDMENT 54-55 (1966).

7. *See* Elonis v. United States, 135 S. Ct. 2001 (2015).

category of speech – true threats – a category in which it has offered virtually no guidance since it was first formally identified by the Court almost half a century ago.

Why not Absolutism?

The most obvious answer to the question "why not absolutism?" is that expression is ubiquitous. Words can defraud, extort, coerce, destroy reputations, and threaten lives. An absolutist view of the First Amendment, in light of the ubiquity of speech in human life, would inject what is quite simply an unacceptable degree of lawlessness into the social world. It would mean the end of contract law, for there could be no adverse legal repercussions for not adhering to contractual terms spoken or written. The words: "your money or your life" would be protected speech. A Neo-Nazi leader could command a follower to "pull the trigger" of a gun pressed firmly against the cheek of an African American – and suffer no consequences. An angry colleague could, without any fear of adverse legal consequence, fabricate accusations of pedophilia and proceed to destroy the career of his targeted victim by spreading word of these falsely manufactured proclivities.

In short, it is virtually impossible to envision a regime of truly absolute free speech. There are many examples, like the hypotheticals above, all of which arguably turn on common sense – circumstances in which most would agree that a First Amendment exception is called for. Of course, these are the easy cases. This leads us to two further complications. First, not all cases are this easy. One person's common sense exception might be another's outrageous incursion into individual freedom. Second, whether addressing a purportedly "easy" case or a more "difficult" one, how are courts to manage this dilemma? Courts are confronted with what seems to be an unenviable requirement that they selectively diverge from a straight-forward reading of constitutional language. How is this to be done?

One's first impulse might be to suggest that courts treat freedom of speech like any other public policy goal. On one hand, we value free speech; on the other, we value other social goods such as enforceable contracts, public safety, and

compensation for those wrongfully harmed. Confronted with circumstances in which free speech conflicts with another important social value, a court could simply balance the interests at stake and arrive at the outcome that best fits the unique set of facts, circumstances, and interests of the case at hand. This is the job of courts after all: to make judgments.

Yet, something about this approach might not sit well. Courts would be tasked with a role virtually indistinguishable from that of their legislative peers: simply weighing the policy interests at stake and choosing the more "socially good" or least "socially harmful" option. Courts would, in this respect, behave just like legislators, drawing upon their personal philosophy, accumulated wisdom, and evidence presented, to arrive at a policy conclusion they think optimal: uninhibited free expression on one side of the scale and counter-veiling interests on the other. Such an approach might be said to blur the important distinction between the legislative and judicial roles.[8] The great First Amendment scholar and consummate critic of ad hoc balancing, Thomas Emerson, lamented that as a doctrinal test, ad hoc balancing "frames the issues in such a broad and undefined way, [and] is in effect so unstructured, that it can hardly be described as a rule of law at all."[9] Even more damning, to Emerson, ad hoc balancing "gives no real meaning to the [F]irst [A]mendment . . . [I]t amounts to no more than a statement that the legislature may restrict expression whenever it finds it reasonable to do so, and that the courts will not restrain the legislature unless that judgment is itself unreasonable."[10]

Furthermore, such balancing might suggest that all policy interests – including those rooted in the Constitution – are on the same plane, but they are not. Article VI of the Constitution declares that the Constitution "shall be the supreme Law of the Land."[11] More than two centuries ago, Chief Justice Marshall scoffed at the notion that ordinary legislation and the Constitution should be treated as equals, emphatically describing as "a proposition too plain to be contested, that the

8. EMERSON, *supra* note 6, at 55.
9. *Id.* at 54.
10. *Id.* at 55.
11. U.S. CONST. art. VI, cl. 2.

constitution controls any legislative act repugnant to it."[12] To Justice Chief Marshall, the implications of adopting an alternative construction would be dire and dispiriting: "then written constitutions [would be no more than] absurd attempts, on the part of the people, to limit a power, in its own nature illimitable."[13] What was intended by the framers to be a rigid constraint on government power – circumscribing its role by identifying particular values that were not to be breached – would instead become just one more policy consideration to be thrown into the mix by both the legislature and judiciary alike. The very fact that the framers chose to include free speech as a fundamental guarantee in the Constitution arguably precludes this approach. The balance has already been stricken. As Justice Black has observed, "a governmental policy of unfettered communication of ideas does entail dangers. To the Founders of this Nation, however, the benefits derived from free expression were worth the risk. They embodied this philosophy in the First Amendment's command . . . ."[14]

A proponent of balancing might retort that balancing need not suggest that all interests be accorded equal weight. A preponderance of the evidence in favor of curtailing free speech – need not be sufficient. In other words, one way to ensure that constitutional provisions like the First Amendment are not reduced to the status of ordinary law – merely by virtue of accepting that they cannot be understood to be "absolute" – is to set the bar for overriding that constitutional interest very high. A range of balancing tests are commonly employed throughout constitutional law, whether or not the Court characterizes its analysis as such. Its approach has not been monolithic. The Court has utilized a variety of standards establishing the requisite interest needed to override a constitutional right. It was the assumption that an easy-to-pass rationality standard would guide ad hoc balancing – with the natural corollary of a deferential judiciary gladly willing to accept a legislative determination that free speech may be abridged – that motivated much of Emerson's ire against this

12. Marbury v. Madison, 5 U.S. 137, 177 (1803).
13. *Id.*
14. Dennis v. United States, 341 U.S. 494, 580 (1951).

doctrinal approach.[15] However, other balancing standards might require much more: a "compelling interest" may be demanded to outweigh free speech.[16] Other circumstances may call for an intermediate analysis, where an "important objective" is required – more than a merely rational interest, but less than a compelling one.[17]

This framework is, of course, familiar to anyone with a rudimentary understanding of Constitutional Law, and is perhaps most commonly associated with the Court's Equal Protection jurisprudence. These tiers of review are also no stranger to the Court's First Amendment jurisprudence. In some First Amendment settings, the Supreme Court has adopted this mode of analysis, and one might conclude that this is our answer to our dilemma. To accommodate the problem of a First Amendment that simply cannot be absolute – as the most ardent free speech advocate may, in theory, wish it could be – we handicap one side of the scale, and we ratchet-up what it required to outweigh free expression.

The Early Years

We might, in fact, characterize the first fifty years of significant First Amendment jurisprudence by the Supreme Court, beginning in the early part of the 20th century, as a progression of this sort of balance-centric thinking. The earliest First Amendment cases utilized the language of balancing. Although in 1919 Justice Holmes introduced what might, on-its-face, have looked like a black-or-white dichotomous rule (the "clear and present danger" test), a cursory read of *Schenck v. United States* reveals a methodology that is much more sliding-scale than on-off switch.[18] A clear and present danger allowing for speech suppression was not to

15. EMERSON, *supra* note 6, at 55.

16. *See* Brown v. Entm't Merc. Ass'n, 564 U.S. 786, 799 (2011) ("Because the Act imposes a restriction on the content of protected speech, it is invalid unless California can demonstrate that it passes strict scrutiny—that is, unless it is justified by a compelling government interest and is narrowly drawn to serve that interest.")

17. *See* United States v. Alvarez, 132 S. Ct. 2537, 2551-2556 (2012) (Breyer, S., concurring).

18. Schenck v. United States, 249 U.S. 47 (1919).

be established by a crystal clear rule that could be applied with uniformity, predictability, and certainty in any setting. Rather, in Justice Holmes' own words, determining whether the standard has been met was "a question of proximity and degree."[19] It is, in short, a balance.

Justice Holmes explained that the First Amendment, quite simply, could not reasonably be understood to be absolute, by way of the metaphor that would become the universal shorthand for this most intractable quandary: "The most stringent protection of free speech would not protect a man in falsely shouting fire in a theatre and causing a panic."[20] The Court's early free speech jurisprudence struggled mightily to identify an appropriate test for managing the stubborn First Amendment problem of an absolute right that could not be absolute. It had a number of options. The Court could have just thrown up its hands and said "we're not touching this issue." For much of the Court's history, it effectively did just this – not unlike the way it had done in 1849 with another discrete provision in the Constitution: the guarantee of a republican form of government in Art. IV, §4.[21]

But with the turn of the 20th century, the Court began to confront the First Amendment dilemma head on. It would not take long for Justice Holmes, after considerable trial and error, to come to the realization that a highly malleable balancing test was not the optimal solution. In his famous dissent in *Abrams v. United States*, he implored the Court to adopt a much more rigorous test – one that was fixed and precise.[22] Justice Holmes proclaimed that even in the context of war, "the principle of the right to free speech is always the same."[23] What he proposed now looked completely unlike the balancing he had utilized in *Schenck* earlier in that very same year.[24] Justice Holmes made it clear that suppression should only be permitted where there was a "present danger of immediate evil

19. *Id.* at 52.

20. *Id.*

21. *See* Luther v. Borden, 48 U.S. 1 (1849).

22. Abrams v. United States, 250 U.S. 616, 624-31 (1919) (Holmes, J., dissenting).

23. *Id.* at 628.

24. *See Schenck,* 249 U.S. at 47.

or an intent to bring it about . . . ."[25] This was no flexible balancing standard. It was a *rule*: one either passed or failed.

Over time, variations of the Holmes formulation – forged in multiple dissents alongside Louis Brandeis – would gain currency on the Court. However, it would take time for the test's precise form to solidify; as a standard, the clear and present danger test would remain in flux for many decades. Justice Jackson quipped: "All agree that it means something very important, but no two seem to agree on what it is."[26] Long after the departure of Justices Holmes and Brandeis, the Court would continue to speak out both sides of its mouth, at once utilizing the firm and unbending "clear and present danger" language of the preeminent justices, but in application imposing it as a malleable balancing test. *Dennis v. United States*, a 1951 red-scare era decision, is perhaps the most notable example of how the Court straddled this line.[27]

In this deeply conflicted – and today, widely condemned and discredited – decision, the Court moved in concert with a country consumed by "Cold War hysteria" that was believed by many to have "infected the judges' reasoning."[28] A plurality of the Court, utilizing an unabashedly balance-based variant of the clear and present danger Test, allowed for the conviction of the major leaders of the Communist Party USA while explicitly citing the Court's movement over time toward a "Holmes-Brandeis rationale."[29] The defendants' crime was neither attempting to nor advocating to overthrow the government, but rather, merely conspiring to advocate such overthrow by organizing the teaching of Marxist-Leninist doctrine.[30] Under the plurality's rationale, the clear and present danger requirement was to be applied as a balancing test that considered not merely the existence or non-existence of imminent danger – what Justice Douglas in dissent would have limited to circumstances in which "conditions are so critical

25. *Abrams*, 250 U.S. at 628.

26. *Dennis*, 341 U.S. at 567 n.9.

27. *Id.* at 494.

28. Christina E. Wells, *Fear and Loathing in Constitutional Decision-Making*, 2005 WIS. L. REV. 115, 118-19 (2005).

29. *Dennis,* 341 U.S. at 507-08.

30. *Id.* at 582 (Douglas, J., dissenting).

that there will be no time to avoid the evil that the speech threatens"[31] – but also the "gravity" of the evil presented. As with any flexible balancing test with multiple variables, increased severity in one area may mean that a lower threshold will suffice in another. Because the gravity of the potential harm was deemed so great – violent overthrow of the government – the fact that the harm was improbable (and certainly not imminent) was not an obstacle to conviction.[32]

In retrospect, there is no question that this approach to the First Amendment is viewed by most scholars, historians, and jurists as most regrettable. Even at the time of the *Dennis* decision, there was deep disagreement on the Court; it split into multiple opinions, and only four of the justices agreed on the use of the particular clear and present danger balancing test formulation that led to the Communists' loss.[33] Justice Frankfurter derided the Court's previous – more categorically-bound – use of the test, while conceding that reinterpreting the test's meaning such that "'clear' and 'present' [now] mean[s] an entertainable 'probability'" was, in light of the Court's precedent, simply inconsistent.[34] In contrast, Justice Jackson, while also arguing that the clear and present danger test should not apply in this situation, was fully comfortable with the test's prior use under alternate circumstances.[35] It would seem that to Justice Jackson, the Communist threat of the 1950s was different in kind from the "trivialities that were being prosecuted" in the earlier part of the century when the test first emerged.[36] "Unless we are to hold our Government captive in a judge-made verbal trap, we must approach the problem of a well-organized, nation-wide conspiracy . . . realistically . . . ."[37]

What is Justice Jackson saying here? A "judge-made verbal trap" is simply a pejorative description of a categorical rule-based approach, one that takes discretion from the legislature

31. *Id.* at 585.
32. *Id.* at 509.
33. *See Dennis*, 341 U.S. 494.
34. *Id.* at 527 (Frankfurter, J., concurring).
35. *Id.* at 567-68 (Jackson, J., concurring).
36. *Id.* at 569.
37. *Id.* at 568-69.

by imposing "either/or" rigidity rather than the fluidity of balancing. While such a methodology may be appropriate in some contexts, to Justice Jackson, this was not one of them. Whether it was conceived of as part of the clear and present danger test, or independent of it, balancing had won the day. However, other impulses were brewing on the Supreme Court. Indeed, less than a decade earlier, in what is perhaps some of the most influential Court dicta in First Amendment jurisprudential history, the Court had laid out a framework for an entirely different approach.

The Rise of the Categorical Approach

There can be no question that seeking to generalize the Supreme Court's evolving doctrinal rules or standards can be a risky endeavor. The Court is a moving target – nine moving targets to be precise. The justices frequently disagree on the appropriate method of resolving constitutional dilemmas. Time, context, and social change can sometimes justify a shifting doctrinal focus, and the Court's membership itself changes – and with it the judicial philosophy of the individual justices. We are often left with the impression that while there may be a dominant doctrinal structure, its boundaries and justifications are disconcertingly blurry. Such is the case with the categorical approach to content-based speech abridgement.

The famous dictum in the 1942 case of *Chaplinsky v. New Hampshire*, more than any other decision, helped solidify the Court's move toward a categorical structure and away from balancing.[38] *Chaplinsky,* unlike many of the Supreme Court's high-profile First Amendment decisions, addressed nothing as heady as a proposed violent overthrow of the government. Instead, the case revolved around a seemingly mundane topic: the spontaneous exchange of angry and aggressive words – akin to what one might imagine would precede a barroom brawl.[39] What was critical to the decision was the Court's broader-than-necessary assessment of how such so-called "fighting words," along with other discrete types of speech,

38. *See* Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).
39. *Id.*

were to be treated for First Amendment purposes. This dictum bears repeating:

> [I]t is well understood that the right of free speech is not absolute at all times and under all circumstances. There are certain well-defined and narrowly limited classes of speech, the prevention and punishment of which have never been thought to raise any Constitutional problem. These include the lewd and obscene, the profane, the libelous, and the insulting or "fighting" words – those which by their very utterance inflict injury or tend to incite an immediate breach of the peace. It has been well observed that such utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.[40]

This much-cited language clearly stood for the baseline proposition that content-based limitations on speech were to be approached categorically rather than through ad hoc balancing. As mentioned previously, for many advocates of a vigorously speech-protective First Amendment, this is a preferred approach. Yet, the dictum itself was subject to harsh criticism from many First Amendment scholars.[41] Although it has endured as precedent establishing this two-tiered method of First Amendment interpretation and as a statement of how and why this doctrinal approach should operate in practice, it was deeply flawed. First, the notion that punishing or prohibiting the enumerated categorically-excepted classes of speech does not "raise *any* Constitutional problem"[42] has never been widely, and certainly not uniformly, accepted. From

40. *Id.* at 571-72.

41. *See* Ronald K.L. Collins, *Exceptional Freedom – The Roberts Court, The First Amendment, and the New Absolutism*, 76 ALB. L. REV. 409, 414-15 (2013).

42. *Chaplinsky*, 315 U.S. at 572 (emphasis added).

fighting words[43] to libel[44] to obscenity[45] – while entitled to a lower level of constitutional scrutiny – the Court has repeatedly concluded that some constitutional protection still applies. Excepted categories are not invisible to the Constitution, as the *Chaplinsky* dictum implies. Second, and perhaps most glaring, is the fact that 40 percent (or two of the five) of the illustrative categories causally laid out as examples of unprotected speech are largely protected.[46] *Chaplinsky* also did not profess to provide an exhaustive list of those categories – leaving many questions for the future. This leads to the third weakness of the *Chaplinsky* dictum: on what bases are the excepted categories to be chosen? Here, Justice Murphy, who authored *Chaplinsky*, provides some guidance. But that guidance is arguably confusing and flawed. Most troubling, to many scholars the dictum emphasized the wrong criteria for determining what categories should fall outside of the First Amendment's full protection.[47]

Chaplinsky conflates two important assertions about how the First Amendment should work: first, that narrowly cabined categories of speech are not entitled to the full First Amendment guarantee (and implicitly that all other speech is entitled to *full* protection) and second, that these categorical exclusions are to be defined, at least in part, by virtue of their low value as speech.[48] Acceptance of the first assertion does not necessarily require acceptance of the second. Yet, this conflation has resulted in decades of doctrinal and normative confusion from prominent scholars and jurists. Some ardently reject the categorical approach, not because there is something inherently objectionable about a two-tiered First Amendment, but because these tiers are seen as inextricably linked to the premise that the First Amendment allows for judicially imposed assessments of the relative value of particular

43. *See* R.A.V. v. City of St. Paul, 505 U.S. 377 (1992).

44. *See* New York Times Co. v. Sullivan, 376 U.S. 254 (1964).

45. *See* Stanley v. Georgia, 394 U.S. 557 (1969).

46. *See*, *e.g.*, Sable Comm. of California, Inc. v. F.C.C., 492 U.S. 115, 131 (1989) (holding that lewd dial-a-porn services are protected expression); Cohen v. California, 403 U.S. 15, 26 (1971) (declaring that profane expression is protected speech).

47. *See infra* pp. 21-27.

48. *See Chaplinsky*, 315 U.S. at 568.

speech.[49] Many scholars see this premise as a perversion of the most foundational principle behind the First Amendment, "that the government has no business evaluating the content of speech."[50] It is anchored in the idea most eloquently articulated by Justice Holmes in *Abrams* that sovereignty of "we the people" was not to be governance by a Platonic guardian who has a superior ability to divine what ideas have value and what ideas do not.[51] The widely-accepted "fighting faiths" of one era, may be "upset" by time, and become the low-value speech of another – or vice versa.[52] Members of the modern Court have, on occasions, explicitly, and in strong language, rejected the value-assessment basis for denying First Amendment protection to particular speech.[53]

There are, however, many potential bases for justifying particular categorical exclusions from full First Amendment protection. While the *Chaplinsky* dictum is widely understood to stand for a categorical approach rooted in an assessment of particular speech's relative social *value*, a close examination of the language reveals a much more nuanced set of possibilities. First, the language could be read to impose a strictly (or predominantly) historic approach to identifying excluded First Amendment categories. Indeed, before Justice Murphy pontificates as to *why* certain categories have been understood to be unprotected, he defines these very categories on the basis of their historically unprotected status, asserting that regulations on these "certain well-defined and narrowly limited classes of speech . . . *have never been thought* to raise any Constitutional problem."[54] In other words, *Chaplinsky* might be understood not as an invitation for the future – to make fresh assessments of individual classes of speech and declare them unprotected – but as a simple statement of historical fact: a certain fixed set of excepted categories exists by virtue of their time-tested status. This approach—focusing on historical

49. *See infra* pp. 21-27.

50. Jeffrey M. Shaman, *The Theory of Low-Value Speech*, 48 SMU L. REV. 297, 300 (1995).

51. *Abrams,* 250 U.S. at 630.

52. *Id.*

53. *See, e.g.*, Meyer v. Grant, 486 U.S. 414, 419 (1988).

54. *Chaplinsky*, 315 U.S. at 571-72 (emphasis added).

tradition as *the* criterion to determine whether or not a particular categorical exclusion exists – has in fact achieved prominence on the Roberts Court.[55] Some scholars have criticized claims that the particular low-value classes of speech articulated in *Chaplinsky*, and others subsequently added by the Court, are truly rooted in history.[56] Nonetheless, regardless of whether one understands the historical consensus on certain speech categories to be largely valid, a strategic exaggeration, or a complete myth, one reading of the *Chaplinsky* dictum would seem to endorse this approach. Furthermore, there might be strong speech-protective instrumental arguments that favor a tradition-based justification for the excluded categories that do exist – most notably, the relative difficulty of adding new unprotected categories under this model.

A second possible reading of the *Chaplinsky* dictum might emphasize that, prior to addressing the purported nominal social value of certain categories of speech, Justice Murphy alludes to the harm imposed by such speech. Recall that in summation, at the conclusion of the sentence that lists the five illustrative unprotected categories, Justice Murphy encapsulates their character as words "which by their very utterance inflict injury or tend to incite an immediate breach of the peace."[57] In the following sentence, Justice Murphy highlights the "social interest in order and morality"[58] that is served by inhibiting such expression. Thus, another quite rational interpretation of *Chaplinsky* – also consistent with its language – would be that the categorical approach is fundamentally about carving out narrow exceptions for classes of expression that impose intolerable harm. In other words, the key criterion for excluding from full First Amendment protection certain categories of speech is the damage caused by such expression. The language alluding to "slight social value"

55. *See* Gregory P. Magarian, *The Marrow of Tradition: The Roberts Court and Categorical First Amendment Speech Exclusions*, 56 WM. & MARY L. REV. 1339, 1341 (2015).

56. Genevieve Lakier, *The Invention of Low-Value Speech,* 128 HARV. L. REV. 2166, 2168 (2015).

57. *Chaplinsky*, 315 U.S. at 572.

58. *Id.*

might simply be understood as reflecting the rather straightforward belief that if something is sufficiently harmful, its potential social value or utility as an "essential part of any exposition of ideas" is negated.[59]

But let us suppose that one does not accept this interpretation – one that would largely reject the independent significance of the low-value language in the *Chaplinsky* dictum in favor of one that stresses only the harm principle. As an alternative, one might assert that harm and social-value are two independent variables – both of which important under *Chaplinsky*. In other words, speech may be harmful, but at the same time have considerable social value, just as completely harmless speech may have the hollow social value of a supermarket tabloid. Admittedly, the dictum could be faulted for being somewhat unclear about the relationship between these two potentially independent attributes of speech – perhaps misleadingly implying that they are necessarily linked. But this ambiguity might be boiled down to poor draftsmanship – and this would not be the first time the Supreme Court has been accused of such a sin.

Thus, a third – and also quite palatable – interpretation of *Chaplinsky* might view the social-value component of its rationale as a variable that is thrown into the mix only where the expression is particularly harmful. In other words, because the harmful categories articulated *also* are of "slight social value," the cost of a deprivation is "outweighed" by the benefit of averting harm. Under this third interpretation, the low-value attribute only becomes relevant in the case of certain harmful speech, and is only relevant as a way of double-checking that the spirit of the First Amendment is not being infringed when an especially harmful category of speech is excluded from full First Amendment protection. If we are confident that this harmful speech we are prohibiting is also "no essential part of any exposition of ideas"[60] and at most, of "slight" value, we can be more comfortable that on balance, allowing the abridgment of such speech is the right thing to do.

However, it is unfortunately a fourth interpretation that

59. *Id.*
60. *Id.*

has dominated the scholarship and the judicial interpretation of *Chaplinsky*'s dictum. Under this interpretation, the categorical approach outlined in the case is foundationally about something called "low-value" speech. As I shall argue, this has led to unmerited criticism of the categorical, or two-level, method of content-based speech protection. Yet, because it has been repeated so frequently, for so many years, this view of *Chaplinsky* lives on. Indeed, First Amendment law school casebooks are sometimes structured around this "low-value" distinction.[61] Adding to the confusion, it is not always clear that when the phrase "low value speech" is used, the author is suggesting that the expression has been deemed less valuable as speech. Sometimes "low value" speech is used as a shorthand catchall for any speech that falls into a less protected category, regardless of *why* it has been placed in that category. It is not uncommon for the words "low value" to be used to describe speech that passed a certain harm threshold or that has simply been historically unprotected.

Where the Scholars Stand: A Sampling

There are, as we have seen, many ways of summing up the doctrinal legacy of *Chaplinsky*. The preeminent First Amendment scholar Ronald K.L. Collins for example, describes *Chaplinsky's* dictum as consisting of just "two separate prongs . . . the categories prong . . . [and] the low-value speech prong [which is] premised less on particular categories of speech than on the value of the expression in question."[62] This summation, while having the virtue of disaggregating the categorical approach from the "low-value" one, does not account for *Chaplinsky's* other possible interpretations, all of which have been reflected at different times in various Supreme Court decisions.[63] The confusion and inconsistency of the

61. *See, e.g.*, GEOFFREY R. STONE et al., CONSTITUTIONAL LAW 1134 (7th ed. 2013).

62. Collins, *supra* note 41, at 417, 422.

63. *See, e.g.*, United States v. Alvarez, 132 S. Ct. 2537, 2551-52 (2012) (Breyer, J., concurring) (advocating for an intermediate scrutiny approach that would look to both the harm and value of the speech and balance their respective weights); Brown v. Entm't Merch. Ass'n, 564 U.S. 786 (2011) (advocating for history and tradition approach); Brandenburg v. Ohio, 395

scholarship in this area is not surprising in light of the Court's lack of clarity.

In a widely cited exchange in Northwestern Law Review, two constitutional law luminaries – Cass Sunstein and Larry Alexander – sparred about the merits of distinguishing between low and high value speech. To Sunstein, valuing speech is something akin to a necessary evil, a "difficult and unpleasant task" that cannot be avoided in an optimal system of free expression.[64] Indeed, Sunstein conceptualizes value classification as a speech-protective alternative to the much-worse prospect, "authoriz[ing] government to distinguish among ideas on their merits."[65] To Sunstein, categorizing certain speech based upon its content where it "lies somewhat afield from the core concerns of the first amendment,"[66] is strongly preferred over raw viewpoint discrimination. And while, according to Sunstein, harm may *also* be a basis for categorizing speech as not fully protected, limiting inquiries to harm alone would be "intolerable."[67] This assessment is ostensibly rooted in speech-protective concerns. Sunstein worries that looking to harm alone would mean opening all speech to harm analysis and, as a consequence, lowering the burden for suppressing speech across the board to all categories of expression.[68] However, Sunstein goes on to acknowledge a concern *not* rooted in speech-protection: that is, if the harm-only approach were interpreted to mean that an equally stringent standard applies to all speech, regardless of whether it is political, commercial, libelous or pornographic.[69] To Sunstein, this approach would "fail to draw lines that ought to be drawn," suggesting that, quite simply, all speech is not created equal.[70] Applying the same singular standard to all speech does not acknowledge this truth. It would also, according to Sunstein, most likely result in tacit but unspoken

U.S. 444 (1969) (advocating for harm-centered approach).

64. Cass R. Sunstein, *Low Value Speech Revisited*, 83 NW. U. L. REV. 555, 557 (1989).

65. *Id.*

66. *Id.* at 555.

67. *Id.* at 558.

68. *Id.*

69. Sunstein, *supra* note 64, at 558.

70. *Id.*

(and unspeakable) value judgments by courts, covertly playing a role in their harm analysis.[71]

Larry Alexander, in contrast, finds no place for "low" and "high" value distinctions.[72] Aside from the practical challenges of parsing the valuable from the less valuable, Alexander sees the distinction as resting on an intractable fallacy: "Such division assume[s] that, for purposes of 'freedom of speech' values, 'speech' resides in an object, such as a printed page, a frame of film, or a series of sounds, rather than in the derivation of meaning from . . . the audience or in the intended meaning of the speaker."[73] Citing just one example, obscenity, he points out that a category of unprotected expression cannot be valued without reference to the context in which it is received or conveyed.[74] Who is to decide whether allegedly pornographic language in a respected novel is to be interpreted in light of the entire work (high value) or in isolation (low value)?[75] What if the pornography at issue is utilized as part of a legitimate academic study on human behavior?[76] For Alexander, the error is not in attempting to distinguish among different types of speech in order to determine, for First Amendment purposes, what can or cannot be restricted, but rather, the use of the value-based taxonomy to accomplish this.[77]

The preeminent First Amendment scholar, Daniel Farber, has likewise rejected the low value distinction, observing that the Court has historically looked to inconsistent sources to classify speech.[78] Yet, Farber argues that with most categories, the Court has been more likely to rely upon a harm principle to justify reduced First Amendment protection. Farber conceptualizes the Court's categorical approach as something like an automatic compelling interest test.[79] The government is

71. *Id.*
72. Larry Alexander, *Low Value Speech*, 83 NW. U. L. REV. 547 (1989).
73. *Id.* at 547.
74. *Id.* at 547-48.
75. *Id.* at 551.
76. *Id.*
77. Alexander, *supra* note 72, at 552.
78. Daniel A. Farber, *The Categorical Approach to Protecting Speech in American Constitutional Law*, 84 IND. L.J. 917, 933 (2009).
79. *Id.* at 932.

presumed to have a compelling interest in restricting certain categories of speech, such that as a default proposition an abridgement of that speech will pass strict scrutiny – whether it is consumer fraud, perjury in a courtroom, or incitement of imminent violence.[80] Although this is not generally the language the Court uses to describe its categorical doctrinal approach, to Farber, it effectively captures what the Court is typically doing.[81] As discussed earlier, there are quite simply some common sense domains that convincingly suggest that the First Amendment cannot be absolute. To Farber, these have typically been limited to categories of speech in which the state – justifiably – has a compelling interest to regulate.[82]

However, not *all* categories have fit this mold. There are anomalous categories, Farber explains. The Court has justified these categories primarily on the basis of their inherent "low value," and not due to a compelling need to suppress.[83] Farber identifies two such categories: obscenity and commercial speech.[84] "Obscenity seems to be proscribed less because it threatens a compelling interest and more because the Court views sexual speech as inherently less valuable than other kinds of speech."[85] Likewise, the Court's commercial speech doctrine is built around a four-part intermediate scrutiny test that explicitly rejects the need for a compelling state interest to regulate such expression; instead, the Court requires just a "substantial interest."[86]

So, other than these categories of obscenity and commercial speech, Farber sees the animating principle behind unprotected categories to be harm-based, a compelling need by the government to regulate certain narrow types of speech. Farber, in other words, inverts a conventional wisdom of many scholars and jurists – rooted in a common reading of the *Chaplinsky* dictum – that the categorical approach is

80. *Id.* at 919.
81. *Id.* at 931.
82. *Id.* at 932.
83. Farber, *supra* note 78, at 933.
84. *Id.*
85. *Id.* at 934.
86. Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n of N.Y., 447 U.S. 557, 564 (1980).

necessarily tethered to a concept of "low value" speech. Instead, he sees the cases where they *are* tethered as anomalous. Indeed, he remains troubled by these low-value-justified categories, expressing concern that the Court's treatment of this expression frequently appears "ad hoc."[87] In the case of obscenity, Farber sees a category that is a mere "reflection of prudish Puritanism."[88] Thus, although Farber appears to be critical of the low value criterion, he characterizes the move toward a categorical approach as "a real judicial achievement."[89]

There is no indication that the categorical approach will be going away any time soon. In recent decades, many members of the Court have struck a similar posture, accepting the two-tiered methodology, but rejecting as its foundation the "low value" criterion.[90] Justice Stevens favorably cited a lower court decision proclaiming that the "First Amendment is a value-free provision . . . ."[91] And it is self-evident why a version of the categorical approach that looks to harm (or the compelling interest of the state), rather than the ostensible value of the speech, is favored by free speech advocates. It would seem to be the narrowest path to resolving the non-absolute absolute problem. Even if, as a practical reality, the First Amendment cannot be absolute, at a minimum it would seem to mandate that governments not be permitted to judge the value of speech for society – or more importantly, for any one individual. Rather, they may only restrict speech where it is absolutely necessary – where there is a compelling interest and no viable alternative.

This harm approach, of course, does come with its own set of risks. Just like the dubious endeavor of determining a universal value of particular speech, a compelling state interest may equally be in the eye of the beholder. In other words, could the harm principle be grease for the descent down a slippery slope? Should every new claim by a government that it has the power to abridge a certain category of speech turn on

87. Farber, *supra* note 78, at 919, 935.
88. *Id.* at 935.
89. *Id.* at 921.
90. *See* United States v. Stevens, 559 U.S. 460 (2010).
91. Meyer v. Grant, 486 U.S. 414, 419 (1988) (citations omitted).

the eloquence of its "compelling" justification? Governments establish new policies every day, and in doing so make new arguments for why these policies are of the utmost importance. Some might fear that the compelling interest model is an open call for new categories of First Amendment exclusion, a recipe for expanding rather than circumscribing speech suppression.

The New History and Tradition Emphasis

Under the leadership of Chief Justice Roberts, the Court's open embrace of the categorical approach has continued to grow in zeal. However, when it comes to defining those categories, its focus has shifted to another aspect of the *Chaplinsky* dictum, and the history and tradition of excluding the category of speech at issue. Gregory Magarian argues that this emphasis is indeed new – akin to "major conceptual innovation" – despite the Court's insistence that by focusing on tradition, it is merely following in the footsteps of the Court's free speech precedents.[92] Magarian may be correct to observe that the reasoning in the Court's recent First Amendment decisions has shifted; however, it is a shift that is arguably rooted in the original *Chaplinsky* dictum – the case that first planted the categorical seed, flawed and imprecise as its language may have been. The setting of this observed shift is also notable. As an area of First Amendment law with an unfortunate legacy of doctrinal ambiguity, the Roberts Court's "categorical" free speech cases were decided in the context of this longstanding need for clarity. The Court was, in other words, simply filling a void. The Court has stressed tradition (or the absence thereof) in a cluster of content-based cases in which it was asked to carve out new categories of unprotected speech,[93] and the Court was in the position of justifying its repeated answer: "No."

Magarian is nonetheless quite critical of the tradition-based analysis that is taking center stage at the same time

92. Gregory P. Magarian, *The Marrow of Tradition: The Roberts Court and Categorical First Amendment Speech Exclusions*, 56 WM. & MARY L. REV. 1339, 1347 (2015).

93. *See Stevens*, 559 U.S. at 460; Brown v. Entm't Merchants Ass'n, 564 U.S. 786 (2011).

that low value and harm-centered justifications fade into the background.[94] Magarian offers a nuanced assessment of a categorical approach that at once lauds the Court for its pro-free speech destination while largely lamenting the route the Court has taken to get there.[95] Of course, history and tradition are hallmarks of contemporary conservative constitutional analysis, and this reliance on the past is hardly limited to the arena of free speech. Thus, some of the very same criticisms that might apply to the use of history and tradition more broadly may apply with equal vigor to the First Amendment. Magarian takes issue with the potential for manipulation of this historical approach, i.e. "the fact that reasonable people disagree both about what traditions exist and about how, and how much, tradition should matter" and whether judges have the "institutional competence" to play the role of historians.[96] One might also lament the way tradition may be used to obscure the underlying substantive reasons for the Court's decision,[97] and stress that limiting speech protection to what was traditionally afforded might ultimately defeat the broader constitutional purpose of preventing *current* political majorities from suppressing *current* minorities.[98]

Regardless of how one generally stands on the heavy reliance on history and tradition so characteristic of today's conservatives, it must be acknowledged that the First Amendment is its own distinctive area of jurisprudence, with its own doctrinal challenges and substantive concerns. The tradition-based orientation has some undeniable benefits (from a free speech perspective) when used in conjunction with the categorical approach. Specifically, it may allay reticence associated with the "low value" and "harm-based" routes to defining unprotected categories. On its face, it would seem to stop both justifications in their tracks. Under the history and tradition construct, there is slim likelihood of new broad declarations that some as-of-yet-undeclared category of speech is not "valuable," and thus unprotected. Enlightened

94. *See* Magarian, *supra* note 55.
95. *Id.*
96. *Id.* at 1356.
97. *Id.*
98. *Id.* at 1357.

modernity just won't cut it under a tradition-based analysis. If misogynistic speech by men-being-men in 1950 was, at the time, a socially-valued display of masculinity, but is today thought to be of "low-value" or even "valueless" by a majority of Americans, such speech would still remain protected; there would simply be no longstanding history or tradition of allowing its suppression.

A similar principle would be true of the slippery slope concerns endemic to a harm-based approach. If it wasn't "compelling" yesterday to suppress a particular category of speech, under the history and tradition approach, it is not enough for it to be "compelling" today. Like a fine wine, history and tradition require patience. The ultimate effect of the history and tradition approach, in other words, is to profoundly limit the ability of courts or policy makers to expand the list of categories to which free speech is not guaranteed. Granted, as the Court has acknowledged, "[m]aybe there are some categories of speech that have been historically unprotected, but have not yet been specifically identified or discussed as such in our case law."[99] However, it would presumably be a daunting task to substantiate such a class, particularly with many decades of vigorous free speech litigation behind us. Presumably, most historically rooted First Amendment exceptions, if they exist at all, have been identified by now. In this respect, the history and tradition approach is likely to be highly speech-protective.

However, these advantages can also cut in the opposite direction. History and tradition might have the effect of locking in the poor categorical choices of the past – perhaps like obscenity or fighting words – even when the justification for unprotected status may appear to be anachronistic to much of the population. A harm or low-value based assessment of categorical exclusions – as opposed to a history and tradition analysis – might mean greater responsiveness to the present. Sexually explicit material that was once considered so deeply offensive that a vast majority of the country "knew it when they saw it," today, may elicit a mere shrug, and to many others even represent a powerful (and valuable) expression of

99. *Stevens*, 559 U.S. at 472.

what it means to be human. Abrasive face-to-face confrontational language that, in an earlier era, appeared harmful as a matter of common sense – because no "self-respecting man" would allow such words to be spoken without resorting to immediate physical retaliation – today, may appear to be an expression of passion that projects the intensity of a genuine human reaction, one that has value in its forceful conveyance of an idea. What were once "fighting words" may today appear, at worst, to be "bad form." One might feel that the speaker could have found a more productive way to express his anger, but not, as in the past, assume that such words will naturally result in a violent, physical response. Our understanding of "value" and "harm" changes over time. A history and tradition approach might prevent us from correcting mistakes from the past – or, at minimum, updating First Amendment doctrine to reflect contemporary reality – and returning a formerly unprotected category to protected status.

Indeed, we might look to the 1971 decision of *Cohen v. California* as an example of such correcting for the past.[100] Almost three decades had passed since the *Chaplinsky* Court casually tossed "the profane" into its list of too-obvious-for-elaboration categorical exceptions to First Amendment protection.[101] Yet, in *Cohen*, the Court unequivocally declared that the display of the words "F*** the Draft" on one's clothing was protected speech.[102] Explicitly relying upon the two-level approach, the Court concluded that "this case cannot be said to fall within those relatively few categories of instances where prior decisions have established the power of government to deal more comprehensively with certain forms of individual expression."[103] It reasoned that this crude language did not, on its own, fall within the established categories of obscenity or fighting words, or of intentionally provoking a group to a hostile reaction.[104] Although the Court did not describe its

100. Cohen v. California, 403 U.S. 15 (1971).

101. *See* Chaplinsky v. New Hampshire, 315 U.S. 568 (1942).

102. *Cohen*, 403 U.S. at 16, 26.

103. *Id.* at 19-20.

104. *Id.* at 20. The only caveat being that, since *Cohen*, the Court has shown some willingness to allow for certain limited restrictions on profanity

decision as such, *Cohen* might be seen as an example of the Court *unaccepting* a class of expression – returning it to fully protected status. In this enlightened period of the early 1970s, the Court's moderately conservative Justice Harlan famously observed – in a manner that was perhaps invisible to a previous generation – that "one man's vulgarity is another's lyric."[105] While the outcome of a historical counterfactual may be stubbornly resistant to proof, we might reasonably conclude that if the Court of 1971 had been as wedded to "history and tradition" as today's Court, it may have been much more reluctant to quietly read "profanity" out of the *Chaplinsky* formulation.

Nevertheless, framing constitutional doctrine is often more about working within the confines of reality than it is about finding perfection. The framers did not provide perfection. Since an absolutist interpretation of the First Amendment is untenable, the Court must struggle to find the next best alternative. Each choice is bound to be fraught with distinct disadvantages, providing unlimited fodder for law review articles and openings for clever First Amendment lawyers. This truth is also what makes choosing not to choose – by, for example, declining to adequately define and delimit established unprotected categories – and simply opting, instead, for a covert form of ad hoc balancing, so tempting. Drawing a bold line according to "history and tradition" may be subject to a range of justifiable criticisms, yet First Amendment scholars may, at the same time, rightfully celebrate the clarity of the line, for this line has amounted to some impressively speech-protective decisions.

It led the Court in 2010 to reject the claim that depictions of animal cruelty should be declared a new category of

similar to traditional "time, place and manner" limitations. These decisions do not permit full scale prohibitions of profane speech; they merely allow narrow restrictions under limited circumstances – such as limiting broadcast on publicly accessible radio that include profanity to times children are less likely to be listening (*see* F.C.C. v. Pacifica Found., 438 U.S. 726 (1978)), or imposing zoning restrictions that limit the location or density of adult establishments. *See* City of Los Angeles v. Alameda Books, Inc., 535 U.S. 425 (2002); City of Renton v. Playtime Theatres, Inc., 475 U.S. 41 (1986); Young v. Am. Mini Theatres, Inc., 427 U.S. 50 (1976).

105. *Cohen*, 403 U.S. at 25.

unprotected speech.[106] The Court heavily rooted its decision in *United States v. Stevens* in the history and tradition rubric, emphasizing the government's failure to identify a "tradition excluding *depictions* of animal cruelty from 'the freedom of speech.'"[107] This, in itself, is notable. But what is arguably even more significant was the Court's seeming rejection of a categorical analysis based in harm and low value – at least in the raw form proposed by the defenders of the law.

As a method of determining whether a new exception to First Amendment coverage should be added, the government proposed "a categorical balancing of the value of the speech against its societal costs."[108] Of course, the *Chaplinsky* dictum does not dictate that its "low value" and "harm" based justifications be used in conjunction with one another, nor that they be employed as part of a balancing test. It could just as easily be understood to provide a strict cut-off: speech with harm above *x* threshold or speech with value below *y* threshold is unprotected by the First Amendment, case closed. Nevertheless, the Court's reliance on a tradition-based analysis allowed it to outright reject the government's formulation, one that would seem to permit an open-ended consideration of a limitless array of newly-proposed categories for First Amendment exclusion under the purported logic that the expression at hand is particularly harmful or valueless. The Court was alarmed by such a suggestion. "As a free-floating test for First Amendment coverage," the *Stevens* majority asserted that the government's proposal was "startling and dangerous."[109]

The Court was making an important point. Conventional wisdom and sheer intuition might suggest that a categorical approach, by cabining those few narrow areas of expression that cannot be offered full protection and guaranteeing the rest, is highly speech protective. However, this is only so if it is structured as a disciplined and circumscribed rule, and not a loose and malleable standard. Calling one's approach "categorical" is not enough to distinguish it from the case-by-

106. *See Stevens*, 559 U.S. at 469.
107. *Id.*
108. *Id.* at 470.
109. *Id.* at 470.

case balancing that many First Amendment advocates decry. Indeed, embracing balancing as part of a two-level methodology might be *more* dangerous to First Amendment values, for unlike case-by-case balancing, a categorical balance speaks to an entire category of speech and has the power to declare vast areas of expression unprotected. The *Stevens* Court does not equivocate on this point:

> The First Amendment's guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits. The First Amendment itself reflects a judgment by the American people that the benefits of its restrictions on the Government outweigh the costs.[110]

What Kind of Categorical Approach?

So, even when we move past the debate over whether or not to employ a categorical approach, we must confront the question of how this categorical approach will work. Should balancing play any role whatsoever in a two-tier First Amendment analysis? Is it ever possible to completely do away with balancing, or is it simply a matter of degree? And what do we mean by "balancing?" Are there alternative methods of balancing that may be used within, or in order to define, particular categories? Although the Court, in the years since *Chaplinsky*, has offered neither clarity nor consistency on these questions, they are matters that have been considered by scholars.

As we saw above, the Court in *Stevens* dismissed the government's suggestion that unprotected categories be determined through a cost benefit analysis, pejoratively referring to this as "ad hoc balancing."[111] However, scholars such as Melville Nimmer have traditionally distinguished the kind of balancing that might be used to define entire unprotected categories of speech from an open-ended balancing

110. *Id.*
111. *Stevens*, 559 U.S. at 460-61.

that would assess each claim of unprotected speech independently.[112] The former, often referred to as "definitional" or "categorical" balancing, was distinguished from "ad hoc" balancing, and was considered more speech protective because it "takes place at a higher level of generalization."[113] Thus, although some balancing might be understood to be essential, to Nimmer, what was critical was that definitional balancing (very much unlike ad hoc balancing) resulted in a clear rule that could be used in future cases.[114] Accepting that such balancing must occur acknowledges a vitally important and difficult role for judges, for they must use their judgment to balance interests.[115] However, with definitional balancing, they are drawing boundaries not only for the present – for the case at hand – but for the future as well.[116] The stakes are, in this sense, much higher than with ad hoc balancing. But these higher stakes serve many speech-protective interests. These stakes may make a court much more reluctant to deny free speech in a particular case because doing so requires setting a categorical precedent that will likely have profound repercussions. A clear rule resulting from definitional balancing also provides greater certainty for both future speakers and future regulators who will act with greater confidence in their expressive (or regulatory) choices without fear that their words will be unprotected or their regulations will be deemed unconstitutional. It reduces the risk of self-censorship by speakers who might have good reason to be insecure; after all, who can predict the outcome of an ad hoc balance of one particular court? It also provides clarity for a government regulator with an interest in combatting a narrow type of especially harmful expression.

Constraints, which come in many forms, are a big piece of what makes a legal system distinguishable from politics. Ad hoc balancing, however, significantly frees judges from

112. See Norman T. Deutsch, *Professor Nimmer Meets Professor Schauer (and Others): An Analysis of "Definitional Balancing" as a Methodology for Determining the "Visible Boundaries of the First Amendment,"* 39 AKRON L. REV. 483, 489-94 (2006).

113. *Id.* at 490.

114. *Id.* at 491.

115. *Id.* at 492.

116. *Id.* at 491.

constraint and gives them the freedom to act more like politicians to be responsive to public pressure, personal predilection, and ideology. As Nimmer pointed out, it is moments of "national hysteria" when the values of free speech become most critical,[117] and when the all-too- natural human impulse for judges to balance away First Amendment rights becomes hardest to resist.[118] As the red-scare-era *Dennis* decision discussed earlier illustrates, accepting ad hoc balancing means giving courts greater leeway to make decisions that, in retrospect, may turn out to be highly regrettable.

Thus, for Nimmer, balancing itself is not the enemy of free speech, as long as that balance is used categorically and not case-by-case.[119] We might note however, that the *Stevens* Court in 2010 collapsed definitional balancing and ad hoc balancing. By referring to the government's proposed approach as "ad hoc balancing," rather than how it would be traditionally labeled – "definitional balancing" – the Court was sending a message that the government's approach to definitional balancing had some of the same free speech dangers traditionally equated with a case-by-case approach.[120] In other words, the government's category-by-category approach looked eerily similar to case-by-case ad hoc balancing. The Court's critique does not come out of thin air.

Alexander Aleinikoff, one of the most prominent scholarly critics of balancing, has argued that even in the categorical setting, balancing should – and can – be avoided.[121] Indeed Aleinikoff – apparently like the contemporary Roberts Court – sees little difference between definitional balancing and ad hoc balancing; as he explained, "[n]ew situations present new interests and different weights for old interests. If these are allowed to re-open the balancing process, then every case becomes one of an 'ad hoc' balance . . . ."[122] To Aleinikoff, the

117. *Id.* at 492.
118. *Id.*
119. *Id.* at 489.
120. *Stevens*, 559 U.S. at 460.
121. T. Alexander Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943 (1987).
122. *Id.* at 980.

widely-held assertion that balancing is inevitable once one recognizes that a particular constitutional right cannot be absolute is simply incorrect.[123] Even though (as discussed) the First Amendment was drafted in absolutist language, categorical exceptions to free speech need not be explained as a result of a balancing of the interests at stake or a cost-benefit analysis. They may be understood "as resting upon a principle internal to the constitutional provision."[124] Yes, the imperative to carve out a free speech exception for revealing military secrets that would imperil the country's likelihood of success during wartime might be explained as a rational balance in which a court decides that the sacrifice of freedom involved is less weighty than the costs of maintaining that freedom. However, to Aleinikoff, such categorical exception could just as easily be explained by way of a principle internal to the First Amendment, that the freedoms within its ambit are premised upon a political system that survives, and functions effectively.[125] Other scholars have pointed to a range of First Amendment values – even though unspecified in the Amendment itself – that may be used to identify excepted categories. These may come from many sources, including the philosophical, historic or political origins underlying the amendment and constitution as a whole.[126] These values, in conjunction with an array of interpretive techniques, may coherently justify particular excluded categories without resort to balancing.[127] Balancing, in other words, even of the definitional variety favored by Nimmer, should not be presumed to be an essential part of a categorical approach to the First Amendment.

Regardless of where one stands one the foundations of the categorical approach to the First Amendment, whether one adopts a low-value, harm-based, or "history and tradition" justification for the categories chosen, or some combination of these three, or whether one believes balancing to be an inevitable part of the categorical approach or something that is

123. *Id.* at 995.
124. *Id.* at 1000.
125. *Id.*
126. Deutsch, *supra* note 112, at 497.
127. *Id.*

avoidable and ideally *should be* avoided, there is also commonality. Once one accepts the categorical methodology as the preferred approach, which many scholars and a majority of the Court have done, it becomes a matter of great interest to assess how it is actually used. It is one thing to say that the categorical or two-level approach will or should be used in the case of a content-based governmental abridgement of speech, but it is another to see how it plays out in practice. It is this question that I move to next. For purposes of this inquiry, I will focus on one particular category chosen by the Court for special non- or lesser-protected First Amendment status: true threats.

Categories, and More Categories

Frederick Schauer has observed that "[t]here are an infinite number of ways of drawing distinctions and therefore an infinite number of potential categories."[128] Since *Chaplinsky,* not only has the number of lesser-protected categories grown, but as the Court's precedents have fleshed out each such category individually, nuances and idiosyncrasies have become evident. It has become increasingly clear that each category serves a distinct purpose.[129] In practice, this has meant that the image of a categorical approach as simple dichotomous, either-or proposition is inapposite. Indeed, today, calling the categorical approach a "two-level" methodology is somewhat misleading. Because "not all forms of speech are necessarily amenable to the same analytic approach,"[130] the Court has developed distinctive methodologies for lesser-protected categories of expression.[131] Even as the categories themselves are relatively static, these methodologies for individual categories have continued to evolve in response to new fact patterns and social understandings that are ever changing.

Three-quarters of a century ago, commercial speech was

128. Frederick Schauer, *Categories and the First Amendment: A Play in Three Acts*, 34 VAND. L. REV. 265, 285 (1981).

129. *Id.*

130. *Id.* at 286.

131. *Id.* at 286-87.

said to receive no First Amendment protection – declared to be an unprotected category just one month after *Chaplinsky* was decided.[132] However, over time, the Court determined that commercial speech was deserving of some protection.[133] Over this same period, the definition of commercial speech was narrowed and specified with greater precision.[134] By 1980, in an opinion drafted by Justice Powell, the Court settled on a four-part test tailored specifically for commercial speech, widely understood to represent a form of intermediate scrutiny.[135] The *Central Hudson* test was the product of a Justice keenly aware of the need for clarity vis-à-vis the categorical method. Justice Powell's law clerk David O. Stewart – who worked on an original draft of the *Central Hudson* opinion – recounted how Justice Powell "worried that the Court's opinions should provide clear guidance to lawyers and judges about how to apply the ruling in future cases. As a lifelong practicing lawyer, he was very sensitive to that concern."[136] In a 1980 memo to Justice Powell, Stewart agreed that now was the time: with commercial speech it seemed "appropriate to try to apply a disciplined approach instead of the more ad-hoc balancing methods used in the earlier cases."[137]

The story was similar with defamation – a type of speech that was listed in the *Chaplinsky* dictum as an example of an unprotected category.[138] Defamation would receive a speech-protective makeover in the celebrated *New York Times v. Sullivan* case of 1964.[139] In *Sullivan*, and a series of cases in

132. *See* Valentine v. Chrestensen, 316 U.S. 52 (1942).

133. *See* Va. State Bd. of Pharmacy v. Va. Citizens Consumer Council, 425 U.S. 748 (1976).

134. *Id.* at 762.

135. Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm'n of N.Y., 447 U.S. 557 (1980).

136. Ronald K.L. Collins, *FAC 6 (First Amendment Conversations) Powell Law Clerk David O. Stewart Discusses the Origins of Central Hudson's 4-Prong Test,* CONCURRING OPINIONS (Jan. 28, 2016), https://concurringopinions.com/archives/2016/01/fac-6-first-amendment-conversations-powell-law-clerk-david-o-stewart-discusses-the-origins-of-central-hudsons-4-prong-test.html.

137. *Id.*

138. Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942).

139. New York Times v. Sullivan, 376 U.S. 254 (1964).

its wake, the Court qualified and dramatically narrowed the types of defamation that may be subject to civil damages, protecting defamatory speech directed against government officials and public figures unless that defamatory falsehood is made with actual malice.[140]

Commercial speech and defamation provide just two illustrations of how the Court might adapt the categorical approach to the complex nature of human expression. Although it may seem desirable to adopt a streamlined two-level methodology in which all non-protected categories of speech are subject to the same analysis, such an approach would come at a high cost. There is often good reason to establish different tests for different categories of speech. Each category presents its own unique set of concerns – distinctive harms and distinctive speech interests. One cannot, for example, fairly assert that the democratic significance of being able to freely criticize public officials – even where the factual foundation for such criticisms prove to be false – is somehow equivalent to the economic effects of making dubious claims about a product in a commercial advertisement. As Schauer points out:

> Freedom of speech need not have any one "essential" feature. It is much more likely a bundle of interrelated principles sharing no common set of necessary and sufficient defining characteristics. It is quite possible that the protection of political discussion and criticism, the aversion to censorship of art, and the desire to retain open inquiry in science and other academic disciplines, for example, are principles not reducible to any one common core.[141]

Threats, like commercial speech and defamation, come with their own distinctive set of concerns. On a common-sense level, threats seem remarkably simple. The straight-forward

140. *See* Curtis Publ'g Co. v. Butts, 388 U.S. 130 (1967); *Sullivan*, 376 U.S. at 254.

141. Schauer, *supra* note 128, at 277.

definition of a "threat" is uncontroversial: "a statement that expresses the speaker's intention to harm the target."[142] As discussed in the beginning of this piece, threats present the quintessential illustration of why we need First Amendment exceptions: it would be the very rare individual who does not see a need to punish the words "your money or your life." At the same time, relative to other excepted categories of speech – such as child pornography, commercial speech and defamation – defining a threat in the real world can be remarkably difficult.

Human expression is deeply contextual. 'Expressing an intention to harm' could include the physical insecurity imposed by an aggressive negotiator who affords little personal space to his listener, and who speaks with sharp language and great intensity. Body language, facial expressions and physical proximity – even without explicitly threatening words – may, to many, convey a threat. On the other hand, threatening words of the most direct variety, which may, if printed on a piece of paper, leave no doubt that a threat was present, might have a completely different meaning when heard in context. The threatening words might have been quoting someone else's speech from long in the past, or might have been spoken by a comedian on a stage. They might convey irony, sarcasm or hyperbole. Furthermore, a threat, unlike child pornography, requires subjective perception. A listener might hear a threat when the speaker only intended to convey excitement. An ineffectual provocateur may do everything within her power to threaten her targeted victims, yet her words may be laughed off by everyone in earshot – taken by no one as a serious threat. And what of the socially tone-deaf listener? She might not feel threatened by words that a reasonable person – in a similar context – would perceive to be a threat. Should this speech be protected? And what of the tone-deaf speaker whose words would reasonably be interpreted as a threat, but who had no intention to threaten, or who lacked knowledge that his expression would likely be interpreted as a threat?

These are difficult questions. With a categorical approach,

142. Kenneth L. Karst, *Threats and Meanings: How the Facts Govern First Amendment Doctrine*, 58 STAN. L. REV. 1337, 1355-56 (2006).

however, difficult questions are inevitable. It may be clear that a categorical exception for a certain type of speech is needed, but this is just the first step. To avoid relapsing into ad hoc balancing, the courts must next move on to the crucial business of getting into the weeds. Distinctions must be parsed, and difficult questions must be answered in a principled manner. This takes courage, because Supreme Court precedents are not easily undone. But a categorical approach without defined categories is arguably worse than a half-measure; over time, it may become a constitutional wound. It says: we will take the profound (but necessary) step of declaring an entire category of speech outside the ambit of a constitutional provision that is, by its own terms, absolute, but we will not finish the job. We will allow lower courts to come to a range of inconsistent conclusions about how that categorical exclusion should work in practice.

With many areas of constitutional law, leaving open unanswered questions is not terribly problematic. However, the concerns are much greater with speech. For the longer the high court sits back and declines to decide, the more uncertainty settles in. The Court's own precedents, in fact, have acknowledged a sort of special status for free speech, requiring greater precision from First Amendment decisions than is required in other areas of constitutional adjudication.[143] As the Court itself has many times acknowledged (and many tyrannical leaders have long understood), legal insecurity is one of the most potent enemies of expressive freedom. Where it is unclear what one may or may not say without legal sanction, where one's fate is in the hands of a court that might – or might not – decide that one's speech is unprotected, where one has little ability to predict which way a court will rule, self-censorship is a likely result.

Fifty years ago, the Court delved into the weeds with regard to the excluded speech-category of defamation. With a keen sensitivity to the First Amendment stakes at hand, the Court rejected an ad hoc case-by-case approach and articulated

143. *See* Gooding v. Wilson, 405 U.S. 518, 520-21 (1972) (establishing that the overbreadth doctrine allows courts to strike down overly broad laws that risk impinging on free speech even where the defendant could constitutionally be penalized).

crisp rules for determining when speech is unprotected defamation and when even defamatory speech is to be shielded by the First Amendment from liability.[144] This allowed for certainty where there was formerly insecurity. It gave the press freedom to aggressively criticize political leaders without the speech-inhibiting fear that an inadvertent factual error would lead to financial ruin. *New York Times v. Sullivan* was, as Alexander Meiklejohn put it, an "occasion for dancing in the streets."[145] When it comes to threats, however, the contrast could not be greater. The Court has distinguished so-called "true threats" from all others – explaining that only "true threats" are unprotected.[146] However, other than providing this initial distinction, the Court's guidance on the true threats category has been despairingly paltry.

The True Threats Category

In 1969, the First Amendment was on an upswing. This was the year *Watts v. U.S.* was decided, the first official acknowledgment of a categorical exception to the First Amendment for true threats.[147] This final year of the Warren Court was an optimistic period for free speech jurisprudence. The Court seemed to understand that an embrace of a categorical approach meant a concomitant responsibility to define with precision just what those categories were. Not only had the Court just decided *New York Times v. Sullivan* five years earlier, clarifying and narrowing the defamation exclusion, in *Brandenburg v. Ohio* it had just resolved the most vexing and longstanding of categorical first amendment issues.[148] Here, it decisively cast off the balancing approach of *Dennis* in favor of a clear and circumscribed definition of incitement.[149] To fall outside of the First Amendment's

144. *See* Gertz v. Robert Welch, Inc. 418 U.S. 323 (1974); Associated Press v. Walker, 388 U.S. 130 (1967); New York Times v. Sullivan, 376 U.S. 254 (1964).

145. Harry Kalven, *The New York Times Case: A Note on "The Central Meaning of the First Amendment,"* 1964 SUP. CT. REV. 191, 221 n.125 (1964).

146. Watts v. United States, 394 U.S. 705, 708 (1969).

147. *Id.*

148. *See* Brandenburg v. Ohio, 395 U.S. 444 (1969).

149. *Id.*

protection, "advocacy of the use of force or of law violation . . . [must be] directed to inciting or producing imminent lawless action and . . . [must be] likely to incite or produce such action."[150] In one fell swoop, the mother lode of categorical dilemmas – one that had plagued the Court for fifty years – was largely resolved. Through a principled and relatively precise articulation of the category of incitement, the Court radically narrowed the freedom of future courts to do what was done in *Dennis*, to bend to the political and social pressures of the times and expand the ambit of an unprotected category at will.

Watts, while not as ambitious as *Brandenburg*, was of the same cloth. The setting was at the time remarkably routine: a public rally at the Washington mall addressing war and peace and police brutality.[151] The ostensibly threatening statement was made by an 18 year-old malcontent who had just been drafted.[152] His words – "If they ever make me carry a rifle the first man I want to get in my sights is L.B.J." – were followed by laughter by both the speaker and the audience.[153] It took just a few short pages for the Court, in a *per curiam* decision, to blithely dismiss the notion that "the kind of political hyperbole indulged in by the petitioner" could be prosecuted as a crime.[154] This was apparently not a difficult decision, judging from its brevity, but it was a significant one. For not only did the Court definitively assert that true threats are an unprotected category of speech – the statute making threats against the president unlawful was declared facially constitutional – it reflected a very particular way of understanding the categorical approach to the First Amendment.[155]

Definitions must be clearly delineated. "What is a threat must be distinguished from what is constitutionally protected speech," said the Court.[156] How is this to be done? The *Watts* Court embraced a decidedly harm-based approach. There was

150. *Id.* at 447.
151. *Watts*, 394 U.S. at 706.
152. *Id.*
153. *Id.* at 706-07.
154. *Id.* at 708.
155. *Id.* at 707-08.
156. *Watts*, 394 U.S. at 707.

no talk of the "low value" of threatening language. Instead, the Court acknowledged the "overwhelming" governmental interest in keeping the president both safe and free to fulfill his duties without "threats of physical violence."[157] At the same time, it stressed that, even in light of this compelling interest, "the commands of the First Amendment [must be kept] clearly in mind."[158] Ultimately, *Watts* did not go beyond telling us what a true threat *is not*: it is not "political hyperbole."[159] However, the case set the stage for a much more in depth articulation of the contours of true threats, and considering the Court's recent track record of taking categorical definitions seriously, it would have been reasonable to surmise that in the coming years, the Court would take up that opportunity, rather than simply letting the lower courts flail about. But then, there was silence . . . decades of silence.

The next significant decision addressing the category of true threats would not come until 2003. Unfortunately, not only did the Court wait far too long to return to the issue, it would do so in a decision that proved deeply unsatisfying to many on all sides of the free speech divide. As in *Watts*, the Court did not ultimately uphold a true threat. Once again, the Court simply told us what a true threat *could not be*. This time, with a mere plurality, the Court concluded that the "*prima facie* evidence provision" in a Virginia criminal law that banned cross burning was inconsistent with the First Amendment because it meant that the burning of a cross alone could suffice for conviction.[160] The plurality reasoned that, while a symbolic cross burning *may* constitute "constitutionally proscribable intimidation" – i.e. a "true threat" – it may just as well have another meaning.[161] "[S]ometimes the cross burning is a statement of ideology, a symbol of group solidarity. It is a ritual used at Klan gatherings, and it is used to represent the Klan itself. Thus, '[b]urning a cross at a political rally would almost certainly be protected expression.'"[162]

157. *Id.*
158. *Id.*
159. *Id.* (citing VA. CODE ANN. § 18.2–423 (1996)).
160. Virginia v. Black, 538 U.S. 343, 364 (2003).
161. *Id.* at 365.
162. *Id.* at 365-66.

On its face, the holding in *Virginia v. Black* was a speech protective one: The symbolic expression at issue was protected by the Constitution. At the same time, a majority of the Court strongly stated – in what was effectively just dicta – that "a State, consistent with the First Amendment, may ban cross burning carried out with the intent to intimidate."[163] This conclusion was reached despite the fact that by a conventional definition, intimidation is not synonymous with a threat. Granted, true "threats" might be said to be, by definition, "intimidating" – particularly if we are, in part, defining a *true* threat as one that would be effective to a reasonable person. A threat that is not "intimidating" would presumably not be a threat harmful enough to merit exclusion from First Amendment protection. However, the reverse would likely not be said to hold true. The Court acknowledges as much, in its thoroughly circular and self-evident assertion that "[i]ntimidation in the constitutionally proscribable sense of the word is a type of true threat, where a speaker directs a threat to a person or group of persons with the intent of placing the victim in fear of bodily harm or death."[164] The Court glosses over what would seem to be a very important and troubling ambiguity when it qualifies the term "intimidation" by "in the constitutionally proscribable sense of the word."[165]

"Intimidating" expression is not necessarily "threatening." Indeed, intimidation is a common expressive tactic used in all sorts of social settings. Intimidation may be a subconscious byproduct of a speaker with great passion or emotion attempting to convey his feelings to a relatively stoic crowd. Heated exchanges over politics, race, sex, or religion are frequently volatile and can quickly become intimidating to a listener. And a particularly sensitive individual may be more likely to feel intimidated than someone with a thicker skin. Threats are, in other words, a subset of intimidating expression. By using the term "intimidate" throughout the opinion as if it were synonymous with "threaten," the Court obscures the important issues raised about the true threats category. The Court's reasoning further muddied an already

163. *Id.* at 343.
164. *Id.* at 360.
165. *Black*, 538 U.S. at 360.

muddy categorical river. This conflation of intimidation and threat caused much confusion.

Furthermore, as Frederick Schauer has pointed out, a threat has traditionally been confined to a face-to-face confrontation in which a single individual – or relatively small group – is "[placed] . . . in reasonable fear for his personal safety (or personal well-being in a larger sense)."[166] However, cross-burnings were historically used to intimidate entire populations and large swaths of racial, religious, and other minorities. Virginia's statute was seemingly drafted to incorporate such broad based intimidation, *as well as* intimidation against a single target. By failing to clarify what types of intimidation of "group[s] of persons" (in the statute's words) qualify as a true threat, the Court arguably threw into question a central assumption about what an unprotected threat looks like. Indeed, one might be tempted to interpret *Virginia v. Black* as an expansion of the true threats categorical exception such that it now includes so-called hate speech directed at minority groups. Yet, the plurality, near the end of the opinion, unequivocally tells us that this is not the case: "It may be true that a cross burning, even at a political rally, arouses a sense of anger or hatred among the vast majority of citizens who see a burning cross. But this sense of anger or hatred is not sufficient to ban all cross burnings."[167] However, proponents of hate speech laws frequently point to its intimidating impact on minority groups, even when such hate speech does not come in the form of an explicit threat. In other words, it is hardly clear that hate speech and intimidation do not largely overlap. Thus, for a number of reasons, the dream that the Court would at last add some clarity to the true threats category were dashed in *Virginia v. Black*.

Yet, on one front, the Court did seem to speak with consistency, a consistency that had the potential to answer a long-debated query about the true threats exception: In order to be excluded from First Amendment protection, does a threat require subjective intent by the speaker? The locus of the dispute in *Black* turned on the law's prima facie presumption

166. Frederick Schauer, *Intentions, Conventions, and the First Amendment: The Case of Cross-Burning*, 2003 SUP. CT. REV. 197, 213 (2003).

167. *Black*, 538 U.S. at 366.

that there was an "*intent* to intimidate" wherever there is a cross burning.[168] The plurality could not countenance this presumption, describing the many alternative communicative *intentions* that might underlie a cross burning.[169] In condemning the constitutionality of the prima facie provision of the Virginia law, the plurality points to the fact that "[i]t does not distinguish between a cross burning done with the purpose of creating anger or resentment and a cross burning done with the purpose of threatening or intimidating a victim."[170] The clear implication is that intentionality is critical. A cross burning intended to create only anger could certainly intimidate observers, but the Court seemed to be suggesting that the subjective purpose needs to be intimidation for such symbolic speech to be unprotected.

Or not. It turned out that in the years following *Black*, scholars and judges would come to inconsistent conclusions as to how *Black* should be interpreted. This confusion was not limited to the relationship between "intimidation" and "true threats;" it extended to the crucial question of subjective intent. Nevertheless, a number of top First Amendment scholars, including Kenneth Karst[171] and Frederick Schauer,[172] agreed that a natural reading of *Black* does suggest that the true threats category is limited to those threats spoken by individuals who intend to threaten. Of course, as we shall discuss, this does not answer an important subsequent question—what kind of intent? Intent might mean that one's purpose was to threaten, but it might also mean that one is merely aware that one's statement may be threatening, even though the words are spoken for another purpose (artistic expression, release of anger, sending a political message . . . etc.). Nonetheless, all of these nuances become irrelevant if we are to reject the conclusion that the true threat exception includes an intentionality requirement by the speaker, as did a majority of Circuit Courts in *Black's* wake.[173]

168. *Id.* at 347-48.
169. *Id.* at 365-66.
170. *Id.* at 366.
171. Karst, *supra* note 142, at 1358.
172. Schauer, *supra* note 166, at 217-18.
173. Elonis v. United States, 135 S. Ct. 2001, 2018 (2015). "Save two,

From the perspective of the harm principle discussed earlier – the premise that categorical exclusions should be justified by a compelling governmental interest to prevent certain intolerable harms – we might conclude that intent should be irrelevant. After all, whether intended or not, the adverse impact of the threatening message on the listener (whether it takes the form of pervasive fear and insecurity, or coercion to take actions that one would otherwise not) will presumably be unaffected. Larry Alexander has argued that it is a mistake to focus on the intention of the speaker in free speech analysis: "Once the First Amendment is triggered by virtue of government's regulatory intention, the analysis should focus on the ultimate harm the government is seeking to avert by interdicting the receipt of a message and the causal mechanism through which receipt of the message leads to the harm."[174]

However, unless one sees the categorical approach as one of pure dichotomy – either the expression lacks free speech "value" (or imposes intolerable harm) or it does not – there would appear to be some critical First Amendment interests at stake in considering the intent of the speaker. While value and harm focus on the speech itself and its impact, it is the speaker himself who produces the expression. An intent-indifferent definition of true threats would risk encouraging self-censorship. Artists and polemicists whose works include language that could be understood to be threatening – whether it is rap music with aggressive lyrics or derisive social commentary of a sardonic comedian – may be chilled. When every word must be measured to avoid the risk of prosecution, and when every irreverent utterance is a source of uncertainty and insecurity, messages that would and should be protected speech may be withheld from the public conversation out of fear.

It is easy to forget that the Court's traditional strict

every Circuit to have considered the issue —11 in total— has held that this provision demands proof only of general intent, which here requires no more than that a defendant knew he transmitted a communication, knew the words used in that communication, and understood the ordinary meaning of those words in the relevant context. *Id.* at 2018 (Thomas, J., dissenting).

174. Larry Alexander, *Free Speech and Speaker's Intent*, 12 CONST. COMMENT. 21, 25 (1995).

scrutiny test involves not just a need for a compelling state interest, but a narrow-tailoring requirement in which the means of achieving that compelling interest must inhibit no-more speech than necessary to address the harm.[175] A nuanced doctrinal rule governing the true threats category might distinguish among various possible sanctions that would apply to true threats, some being more or less speech-inhibiting and some more or less targeted at remedying the harm imposed. The most urgent, and perhaps effective, method of mitigating the distinctive harm caused by threatening speech may be to simply return a sense of security to the threatened target by way of a temporary restraining order. A higher level of intent might be required to ratchet-up the sanctions imposed. Without the requisite level of intent – wherever it is determined to lie – the First Amendment might only permit threatening speech to result in the imposition of a temporary injunction or compensatory damages, but not criminal penalties or punitive damages. These are, of course, just some possibilities as to how a doctrinal test for true threats might be structured. However, complete intent-indifference, because of its speech chilling potential, might be said to fail the narrow-tailoring requirement.

When the Court chooses to refine, in a principled manner, a category itself and the doctrinal mechanisms that govern it – as it did with defamation in *New York Times v. Sullivan* – the Court shows appropriate sensitivity to the First Amendment interests at stake. The Court acknowledges that identifying the category is just the first step. The mere existence of an excepted category is not an excuse for black or white constitutional adjudication, nor for refusing to sort through confusion and inconsistency at the Circuit Court level. Rather than a blunt instrument that removes large swaths of speech from protection without regard to the subtle distinctions within that category, principled categorical tailoring responds to the unique nature of the excepted category, minimizing the harm to free expression while preserving the important governmental tools essential to prevent intolerable harms. A well-functioning categorical approach performs this function

175. Brown v. Entm't Merch. Ass'n, 564 U.S. 786, 799 (2011).

not on an ad hoc, case-by-case basis, but according to a preemptive tailoring of the category at issue. Consistency and predictability is established by a robust body of precedent that forms a clear set of rules for that category, and the Supreme Court, due to its position at the top of the American judicial hierarchy, is the one body with the capability to do this in a uniform fashion.

The Choice of Categorical Non-Definition: Elonis v. United States

Unfortunately, the Court's courage in taking on the nuanced doctrinal work of parsing an important categorical free speech exception in *Sullivan* stands in stark contrast with last year's *Elonis v. United States*. The case was a highly-anticipated opportunity to, at last, put some meat on the bones of the sparsely defined true threats category; it was also closely watched by scholars and the mainstream press because it was the Court's "first examination of the limits of free speech on social media."[176]

At issue was a federal law making the interstate transmission of "any communication containing any threat" criminal.[177] The law was applied to the lurid Facebook postings of a self-described aspiring rap musician, directed at his estranged wife and others.[178] Lyrics posted "included graphically violent language and imagery . . . often interspersed with disclaimers that the lyrics were 'fictitious,' with no intentional 'resemblance to real persons.'"[179] Some examples of the ominous messages include: "Fold up your [protection-from-abuse order] and put it in your pocket . . . Is it thick enough to stop a bullet?[,]"[180] "I'm not going to rest until your body is a mess, soaked in blood and dying from all the

176. Robert Barnes, *Supreme Court Case Tests the Limits of Free Speech on Facebook and Other Social Media,* WASH. POST (Nov. 23, 2014), https://www.washingtonpost.com/national/supreme-court-case-tests-the-limits-of-free-speech-on-facebook-and-other-social-media/2014/11/23/9e54dbd8-6f67-11e4-ad12-3734c461eab6_story.html.

177. *Elonis*, 135 S. Ct. at 2002.

178. *Id.* at 2004-06.

179. *Id.* at 2005.

180. *Id.* at 2006.

little cuts[,]"[181] and "[e]nough elementary schools in a ten mile radius to initiate the most heinous school shooting ever imagined [a]nd hell hath no fury like a crazy man in a Kindergarten class."[182] The defendant described the postings as "therapeutic"[183] and argued that "[a]rt is about pushing limits."[184] Elonis was nonetheless convicted.[185] The Third Circuit rejected the defendant's contention that, under the First Amendment, a speaker must intend to communicate a threat.[186] It instead agreed with the lower court that a mere intent to express a message that would objectively be understood as a threat would suffice – a reasonable person test.[187]

The Supreme Court did not agree with the courts below. However, it studiously avoided the First Amendment issue, instead opting to decide the case on statutory grounds.[188] And even as a statutory decision, the Court strove to be as parsimonious as possible, holding merely that, here, negligence was not sufficient for criminal liability, but going no further.[189] The statute was silent as to mental state, but the Court reasoned that criminal culpability has traditionally been understood to include some element of conscious wrongdoing.[190] The prerequisite of a guilty mind would thus be read into the law.

This, of course, tells us nothing about whether or not the law would have been *constitutional* had the law explicitly included negligent communications in its definition of criminal threats. Nor does it tell us whether it would be constitutional to penalize expression that is reckless as to whether or not the communication will be interpreted as a threat. If recklessness will not suffice, it also remains unclear whether the threat needs to be purposeful or whether mere knowledge of its

181. *Id.* at 2016.
182. *Elonis*, 135 S. Ct at 2006.
183. *Id.* at 2005.
184. *Id.* at 2006.
185. *Id.* at 2007.
186. *Id.*
187. *Elonis*, 135 S. Ct at 2007.
188. *Id.* at 2012.
189. *Id.* at 2003.
190. *Id.* at 2009.

threatening nature will satisfy First Amendment scrutiny. Other unanswered questions about the true threats category abound: Would the First Amendment calculus differ if this were a mere civil action, rather than a criminal one? Is the type of penalty imposed relevant to the constitutionality of laws criminalizing a true threat? After decades of virtual silence in this critical area, *Elonis* left us with no additional insight into how "true threats" are to be treated as a "low value" category of content based speech under the First Amendment. According to the Court, it was simply "not necessary to consider any First Amendment issues."[191] We are left in the dark.

To Justice Alito – writing in partial dissent – the majority was abdicating its duty. He lamented, "this case is certain to cause confusion and serious problems . . . While this Court has the luxury of choosing its docket, lower courts and juries are not so fortunate. They must actually decide cases, and this means applying a standard."[192] Justice Alito thus proceeds to decide what the Court refused to decide. In doing so, he unwittingly exposes just how doctrinally under-theorized and woefully neglected the true threats category has been, and remains.

In less than three pages, Justice Alito attempts to resolve what is a remarkably consequential question of First Amendment law: whether reckless expression of a threat constitutes a constitutionally proscribable "true threat."[193] He answers in the affirmative.[194] Troublingly, he grounds his analysis in the unresolved and muddled *Chaplinsky* reasoning discussed above. Citing just three Supreme Court cases – none of which, by the way, upheld a "true threat" conviction – Justice Alito recounts the clearly "settled" rule that true threats are not protected by the First Amendment.[195] He explains that "there are good reasons for that rule: True threats inflict great harm and have little if any social value."[196]

191. *Id.* at 2012.
192. *Id.* at 2013-14 (Alito, J., dissenting).
193. *Elonis*, 135 S. Ct at 2014-16.
194. *Id.* at 2016.
195. *Id.*
196. *Id.*

However, just under the surface of what might seem like a truism is a statement that begs the most vital questions and is reflective of the still-pervasive confusion left behind in *Chaplinsky*'s wake. A threat certainly *may* "inflict great harm" and "have little if any social value,"[197] but what if it imposes *some* but not *great* harm? What if the threat *does* have social value – such as a violent threat made against a high profile politician intended to bring attention to, for example, that politician's indifference to the violence suffered by certain segments of the population? One might object to the method of getting the point across, but it would certainly constitute political speech – traditionally considered to be among the highest value expression.[198]

One interpretation then, of Justice Alito's brief matter-of-fact assertion as to why true threats are not deserving of protection, might be that he is putting forth a speech-protective threshold definition for an unprotected "true" threat. In other words, he could be suggesting that all other threats – those that do not inflict great harm or have little if any social value – are necessarily entitled to full First Amendment protection. But if this is his thesis, do *both* the (great) harm attribute and the low value attribute have to be present for a threat to be a "true threat," or will one or the other suffice? It is unclear. In the alternative, these variables could be said to interact, such that to be unprotected by the First Amendment a threat imposing extreme harm has a lower threshold to meet when it comes to an absence of social value, and vice versa.

Of course, a court cannot, and should not, be expected to suss out all of these nuances in a single fact-specific case. Perhaps Justice Alito did not intend his words to constitute a definitional test at all. Perhaps he is suggesting that these attributes are simply self-evident characteristics of threatening speech. However, it does not take a lofty thought experiment to conclude that not all threatening expression is created equal. A friend's threat to retract a dinner invitation may not be hyperbole, and may even be mean-spirited, but it is unlikely to be harmful enough to constitute unprotected speech. Not all

197. *Id.*

198. *See, e.g.,* McIntyre v. Ohio Election Comm'n, 514 U.S. 334, 346 (1995).

threats "inflict," or even risk inflicting, "great harm."

A similar point might be made about the purported low value of threats. Justice Alito attempts to wriggle out of the "low value" difficulty by conceding: "It is true that a communication containing a threat may include other statements that have value and are entitled to protection. But that does not justify constitutional protection for the threat itself."[199] This, of course, is a straw man argument. The true First Amendment conundrum is dealing not with the valuable nature of other expression that might accompany a threat, but with circumstances in which *the threat itself* has significant social value. Even if we ultimately resolve that the harm imposed justifies denying the threat to First Amendment protection, courts should be upfront about the expressive costs of doing so. If the Court is honest about the complexities involved, and the stakes on all sides, it might incorporate these variables into a distinctive "true threats" doctrinal rule, a rule that can be applied consistently and predictably rather than on an ad hoc basis.

We need only look to the Court's defamation jurisprudence to see that this goal is realistically attainable. Instead of shying away from the inherent challenges of a categorical system for content-based discrimination, with defamation, the Court embraced the First Amendment puzzle and incorporated significant nuance into its approach. It attached an array of standards to the categorical non-protection of defamatory speech, differentiating among expression directed at public officials,[200] public figures[201] and private individuals,[202] speech containing intentional falsehoods, merely negligent ones[203] or falsehood where there is no fault at all,[204] speech addressing

199. *Elonis*, 135 S. Ct. at 2016.

200. New York Times v. Sullivan, 376 U.S. 254 (1964).

201. Curtis Pub. Co. v. Butts, 388 U.S. 130 (1967).

202. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345-46 (1974) (concluding that "[s]tates should retain substantial latitude in their efforts to enforce a legal remedy for defamatory falsehood injurious to the reputation of a private individual").

203. *Sullivan,* 376 U.S. at 254.

204. *Gertz*, 418 U.S. at 347 ("[S]o long as they do not impose liability without fault, the States may define for themselves the appropriate standard of liability for a publisher or broadcaster of defamatory falsehood injurious to

matters of public concern versus private matters,[205] and penalties that are punitive versus merely compensatory.[206] Granted, it took time for the Court to flesh out these rules. And of course, the job is never completely done. This is the nature of a common law system in which courts are tasked with resolving specific cases or controversies rather than issuing broad policy prescriptions. It took time for the doctrinal rules governing this excepted category of speech to take shape – and many would argue that the resolution was not perfect. But the Court wisely appreciated its important obligation to concretize this excepted First Amendment category. This is quite a contrast with the true threats category.

Reinvent the Wheel? Well . . . yes

One response might be to suggest a somewhat simple solution. Rather than fretting about the Court's unwillingness to adequately operationalize particular categories of unprotected speech, why not simply utilize one of the templates already devised? With some categories – whether it be commercial speech, obscenity, incitement, or defamation – the Court has indeed articulated detailed definitional tests and nuanced doctrinal rules for applying a categorical First Amendment exclusion. One commentator, in one of the first scholarly law review pieces to address the *Elonis* (non)decision, suggests this very possibility.[207]

Michael Pierce, focusing specifically on the online context, proposes a rule for the true threats category that would borrow the public figure/private figure distinction from the Court's defamation jurisprudence.[208] As with defamation, the requisite

a private individual.").

205. Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc., 472 U.S. 749, 761 (1985) ("In light of the reduced constitutional value of speech involving no matters of public concern, we hold that the state interest adequately supports awards of presumed and punitive damages – even absent a showing of 'actual malice.'").

206. *Id.* at 763.

207. *See* Michael Pierce, *Prosecuting Online Threats After Elonis*, 110 NW. U. L. REV. 51 (2015).

208. *Id.* at 59.

level of intent for a particular threat to be unprotected would depend upon who is targeted by the online threat.[209] The required *mens rea* for a non-protected true threat against a public figure would be a higher (specific intent) than when the threat is aimed at a private individual (recklessness).[210] To Pierce, the same rationale in the defamation context for this two tiered intent standard would justify its use in the setting of online threats.[211] Namely, "the target's identity can serve as a useful proxy for whether the speech attacking him or her has broader significance."[212]

We might rightfully applaud the author's acknowledgement that some threats are bound to have "value" as speech, and that such value will vary from statement to statement.[213] Certainly, threatening language directed at public figures may also communicate a message relating to an issue of public concern. If the public figure is also a public *official* with political duties, it may be even more likely that punishing such threats will, in some sense, stifle public debate and discussion about important issues. A heightened threshold for intent was famously justified in *New York Times v. Sullivan* as a critical speech protective measure because, as the Court explained, "erroneous statement is inevitable in free debate" and it must thus be offered some First Amendment protection if free expression is to have necessary "breathing space."[214] Because of the significant role they play in public life, this need for breathing space becomes more important when criticisms are directed at public figures. There is an unmistakable tendency to discuss important public issues through the lens of such individuals. One might reasonably agree that breathing room for some falsehoods is analogous to breathing room for some threatening language used against pubic figures. In other words, the First Amendment protects the occasional careless factual error about a public figure the same way it should protect the occasional use of inflammatory rhetoric that

209. *Id.*
210. *Id.* at 51.
211. *Id.*
212. Pierce, *supra* note 207, at 54.
213. *Id.* at 51.
214. *Sullivan*, 376 U.S. at 271-72.

might sound like a threat to an objective listener.

However, in other important respects, the public/private figure distinction is not a good fit for the true threats category. Public figure status of a target is a rather crude proxy for threatening speech's value. One might argue, as Pierce does,[215] that the voluntary exposure to risk of defamation is quite comparable to the voluntary exposure to threats one might reasonably anticipate as a public figure. This may be true. A threat against a public figure certainly *may* have significant social value as expression, and it may, perhaps, be more likely, on average, to have such value than threats made against private individuals. But that does not make the analogy to defamation a well-tailored fit for First Amendment purposes. In sharp contrast with defamatory speech, there is no reason to assume that there is a correlation in any particular case between the target of a threat and that threat's social value.

The Court famously concluded in *R.A.V. v. St. Paul* that even within an unprotected category of speech, discriminatory distinctions that prohibit some, but not all, speech are not permissible unless they directly relate the very reason the category is proscribable.[216] This logic fits quite well with the Court's modified test for defamation of public figures; it makes distinctions that relate directly to the reason why defamation is an unprotected category in the first place.[217] *Public* figures, by definition, possess a reputation that is a matter of *public* concern. Reputation is central to the life of a public figure in a way that is qualitatively different than for a private individual. The need to actively manage one's reputation, the expectation that one's affairs will become a matter of public concern, and the greater access to the media and other outlets that facilitate effective use of counter speech to rebut falsehoods, are all part and parcel of the public figure identity. The doctrinal distinction within the defamation category crafted by the Court speaks directly to these qualities. As the Court explained in

215. *See generally* Pierce, *supra* note 207.

216. R.A.V. v. City of St. Paul, 505 U.S. 377, 387-88 (1992).

217. *Defame*, MERRIAM WEBSTER, http://www.merriam-webster.com/dictionary/defame (last visited Nov. 5, 2016) (defining de*fame* as "to hurt the reputation of (someone or something) especially by saying things that are false or unfair").

Gertz v. Robert Welch, public figures voluntarily "have assumed roles of especial prominence in the affairs of society . . . [and] thrust themselves to the forefront of particular public controversies in order to influence the resolution of the issues involved."[218] Public figures and private individuals are quite simply not similarly situated with regard to the harm imposed by the expressive category, nor to the speech value of the expression.

In contrast, increased likelihood of being threatened is merely *associated* with public figure status. This is not unlike the way other qualities might make one more susceptible to being a target of a threat such as wealth, power, a violent temperament or a sharp tongue. Yet, there is little apparent reason to believe that these attributes, like public figure status, should result in a differential First Amendment standard for determining whether *someone else's* speech should be protected. A threat imposes a fundamentally different kind of harm than defamation. A threat is aimed directly at a target. Indeed, the harm imposed by a threat is, by definition, only harmful when the threatening idea reaches the consciousness of the target. The words themselves place the target in a state of fear of an impending concrete physical or psychological injury and may be coercive. Defamation, in contrast, imposes a harm on the target by disseminating false ideas into the greater marketplace. Defamation wrongly diminishes or distorts the perception others have of the target. The harmful ideas do not need to reach the target for the injury to occur. If we take a close look at the doctrinal standard established by the Court for the defamation category, we see that it was thoughtfully tailored to the nature of the particular harm imposed by defamatory speech. True threats are in need of their own similarly fine-tuned test and definition.

Conclusion

The goal of this article is not to propose a new model framework for the doctrinal mechanics of the true threats exception to the First Amendment. Ultimately it is up to the

218. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974).

courts to establish this blueprint, to promote clarity, to further First Amendment values while keeping in mind the unfortunate truth that free speech cannot be absolute. This is the Court's job. With great doctrinal choice comes great doctrinal responsibility. We have seen how the Court made one such choice: the categorical model itself was one way of addressing an intractable problem built into the First Amendment. As with much Constitutional interpretation, the process of refining and rationalizing a particular doctrinal choice takes time. We have seen how the case that first laid the groundwork for the categorical approach, *Chaplinsky v. New Hampshire*, was riddled with ambiguity, leaving many questions and uncertainties for the future. This article has examined how many of these uncertainties stubbornly persist, with members of the Court and top scholars alike still in disagreement as to both how the categorical approach should function and the basis on which particular categories should be identified.

For the foreseeable future, the categorical approach is here to stay. But a categorical system, without adequately defined and systematized categories, is no system at all. It is arguably much less protective than *ad hoc* balancing, for it establishes a default rule that a particular category of speech is unprotected. Without defining that category with precision, or establishing a nuanced test that is tailored to the particular harms and expressive interests unique to that speech category, the Court invites blunt speech-suppressive legislation, uncertainty by speakers whose expression might be chilled, and inconsistency in the courts below. The First Amendment, as written, offers no guidance as to how exceptions should be made to comfortably coexist with its absolutist language. We have seen how well the categorical model can work when adequately fleshed out, as a fair and predicable doctrinal rule. There is every reason to expect the Court to aspire for such clarity wherever gaps in the categorical system exist.